



# Identification, biochemical characterization, and evolution of the *Rhizopus oryzae* 99–880 polygalacturonase gene family<sup>☆</sup>

Jeffrey A. Mertens<sup>a,\*</sup>, Ryan C. Burdick<sup>a,1</sup>, Alejandro P. Rooney<sup>b</sup>

<sup>a</sup> Fermentation Biotechnology Research Unit, National Center for Agricultural Utilization Research, Agricultural Research Service, U.S. Department of Agriculture, 1815 North University Street, Peoria, IL 61604, USA

<sup>b</sup> Microbial Genomics and Bioprocessing Research Unit, National Center for Agricultural Utilization Research, Agricultural Research Service, U.S. Department of Agriculture, 1815 North University Street, Peoria, IL 61604, USA

## ARTICLE INFO

### Article history:

Received 21 July 2008

Accepted 30 September 2008

Available online 4 October 2008

### Index Descriptors:

Polygalacturonase

Family 28 glycosyl hydrolase

*Rhizopus oryzae*

Recombinant protein expression

*Pichia pastoris*

## ABSTRACT

A search of the recently sequenced *Rhizopus oryzae* strain 99–880 genome database uncovered 18 putative polygalacturonase genes with two genes being identical and only one with similarity to a previously reported *R. oryzae* polygalacturonase gene. The 17 different genes share 50% to greater than 90% identity at the nucleotide level as well as the deduced protein sequence level. The cDNA of the different genes was isolated directly or recombinantly and used to express the encoded proteins in *Pichia pastoris*. Recombinant protein expression demonstrated that 15 of the 17 genes encode active enzymes with twelve genes encoding for endo-polygalacturonase enzymes and three genes encoding for exo-polygalacturonase enzymes. Phylogenetic analysis indicates that the genes form a distinct monophyletic group among fungal polygalacturonase enzymes. Finally, our results also suggest that the ancestral form of polygalacturonase in fungi is endolytic and exolytic function evolved later, at least two independent times.

Published by Elsevier Inc.

## 1. Introduction

Endo-polygalacturonase (E.C.3.2.1.15) and exo-polygalacturonase (E.C. 3.2.1.67) along with exo-poly- $\alpha$ -galacturonosidase, rhamnogalacturonase, and endo-xylogalacturonan hydrolase, are classified as members of glycoside hydrolase (GH) family GH28 (Markovič and Janaček, 2001; Niture, 2008). The GH28 family enzymes work in concert with additional glycoside hydrolase enzymes to catalyze the hydrolysis of pectin, a complex carbohydrate structure found in the cell walls and middle lamella of plants (Niture, 2008). Both endo- and exo-polygalacturonase hydrolyze  $\alpha$ -1,4 linked galacturonic acid residues of the homogalacturonan chains of the “smooth” region of pectin with endo-polygalacturonases hydrolyzing within the homogalacturonan chain and exo-polygalacturonases hydrolyzing D-galacturonic acid from the non-reducing end of the galacturonan chain with inversion of the anomeric configuration (Henrissat and Davies, 1997). Due to the efficient hydrolysis of the homogalacturonan chains and demonstrated secretion by pathogenic fungal organisms, polygalacturo-

nase (PG) enzymes are implicated in the invasion and maceration of plant tissues by fungi and as a result are a target for selective inhibition to prevent destruction of economically important crops (Stotz et al., 2000; Saito et al., 2004a; Niture, 2008). Despite the desire to inhibit these enzymes in agricultural applications, the efficiency in which these enzymes macerate tissue is exploited industrially in the clarification of juices and the retting of flax (Henriksson et al., 1999; Zhang et al., 2005; Niture, 2008).

Numerous PG genes, and the enzymes they encode, have been isolated and characterized from a number of prokaryotic and eukaryotic species (Markovič and Janaček, 2001). Most species have been shown to have a small number of PG genes; however, some fungal species have been shown to possess a large family of PG genes with varying biochemical characteristics. The Ascomycete fungal species *Aspergillus niger*, which has as many as eleven PG genes (de Vries and Visser, 2001; Pel et al., 2007), *Botrytis cinerea*, which has six PG genes (Kars et al., 2005), and the basidiomycete *Chondrostereum purpureum*, which has five PG genes (Williams et al., 2002) are perhaps the most well characterized genetically and biochemically (Kester and Visser, 1990; Pagès et al., 2000; Parenicová et al., 2000a; Parenicová et al., 2000b), with the protein crystal structure solved for a number of *Aspergillus* endo-PG enzymes (van Santen et al., 1999; Cho et al., 2001; van Pouderooyen et al., 2003). In contrast to the detailed characterization of the large family of PG genes found in Ascomycete and Basidiomycete species, little information is available on the repertoire of PG genes in *Rhizopus*, a member of the order Mucorales within the fungal

<sup>☆</sup> The mention of firm names or trade products does not imply that they are endorsed or recommended by the U.S. Department of Agriculture over other firms or similar products not mentioned.

\* Corresponding author. Fax: +1 309 681 6427.

E-mail address: [jeffrey.mertens@ars.usda.gov](mailto:jeffrey.mertens@ars.usda.gov) (J.A. Mertens).

<sup>1</sup> Present address: HIV Drug Resistance Program, National Cancer Institute (NCI)-Frederick, Building 535 Frederick, MD 21702, USA.

subphylum Mucoromycotina. To date only one PG gene, *rpg1*, has been isolated from a *Rhizopus* species (Yoshida et al., 2004). This gene, which was isolated from *Rhizopus oryzae*, was classified as an exo-PG due to the deduced protein's sequence similarity to *A. niger* exo-PG enzymes (Yoshida et al., 2004). There is also limited biochemical information on *Rhizopus* PG's with a few reports of a single endo-polygalacturonase enzyme isolated from crude extracts of other *R. oryzae* strains (Saito et al., 2004a; Zhang et al., 2005).

In this work, we report on the characteristics and initial biochemical characterization of 18 PG genes uncovered in a search of the recently released *R. oryzae* strain 99–880 genome ([http://www.broad.mit.edu/annotation/genome/rhizopus\\_oryzae/Multi-Home.html](http://www.broad.mit.edu/annotation/genome/rhizopus_oryzae/Multi-Home.html)). Of the 18 polygalacturonase genes identified in this strain, only one of the genes is nearly identical to the previously isolated *rpg1* gene. Two genes are identical, however the remainder, while highly similar, are unique genes in this species. The cDNAs of the 17 different genes were isolated to confirm the sequence, placement of the introns, and used as the template for expression of the recombinant protein in *Pichia pastoris*. Initial biochemical characterization of the seventeen recombinantly expressed proteins demonstrates that fifteen genes code for active PG enzymes, with twelve found to be endo-PG enzymes and three found to be exo-PG enzymes.

## 2. Materials and methods

### 2.1. Strains and media

Media components and chemicals were obtained from Sigma Chemical (St. Louis, MO) unless otherwise noted. *R. oryzae* strain 99–880, deposited in the Genetic Stock Center as FGSC 9543, was the source of DNA and RNA in obtaining the PG cDNA. Shake flask cultures of strain 99–880 were grown in 50 ml minimal RZ1 media (Skory, 2002) and the desired carbon source at 1% (w/v). Cultures were started by inoculation of approximately  $10^6$  spores and then incubated at 30 °C, 200 RPM. The membrane transfer culture (MTC) was patterned after the method of Ishida et al. (1998). Briefly, a 0.45 micron uncharged nylon (Schleicher & Schuell; Keene, NH) filter was placed on a minimal RZ1 agar plate containing 50% (w/v) maltose to give the effect of low *Aw*, or low moisture content, and the desired carbon source. *R. oryzae* 99–880 spores were spread onto the nylon filter and incubated at 30 °C for 48 h. *R. oryzae* strain *pyr17* (Skory, 2002), a *ura*-strain, was used as the host to obtain cDNA for genes not obtained directly from strain 99–880. Strain NRRL 395 was obtained from the Agricultural Research Service Culture Collection (<http://nrrl.ncaur.usda.gov>). *R. oryzae* 395 was grown in submerged cultures using the same medium as for strain 99–880 with glucose as the carbon source. *Escherichia coli* strain TOP10 (Invitrogen; Carlsbad, CA) was used for construction and propagation of cDNA clones. The wild type *Pichia pastoris* strain X-33 (Invitrogen; Carlsbad, CA) was used for protein expression studies.

### 2.2. Isolation of PG cDNA

The putative PG cDNA's were isolated from total RNA or via recombinant methods. Total RNA was isolated from 50-ml shake flask cultures of *R. oryzae* strain 99–880 grown for 24 h, as above, with either, 0.5% pectin, 0.5% polygalacturonic acid (PGA), 0.5% sucrose, 0.5% glucose, 0.5% xylose, or 10 mM D-galacturonic acid as the carbon source. Total RNA was also isolated from MTC plates containing 0.5% pectin, 0.5% PGA, or 0.5% glucose. Mycelia were disrupted using glass beads and the total RNA was purified using the RNeasy Total RNA kit (Qiagen; Valencia, CA) followed by DNase treatment with the Turbo DNA-free kit (Ambion; Austin, TX) for 30 min. Reverse Transcription (RT) was performed using the Super-

script III First Strand Synthesis System (Invitrogen; Carlsbad, CA) according to manufacturer's instruction using 500 ng total RNA and oligo dT to prime reverse transcription. First strand cDNA was amplified using *Pfu* ultra II Fusion HS DNA polymerase (Stratagene; LaJolla, CA) with gene specific forward and reverse primers (Supplementary Table 1). The forward primer was designed to eliminate the putative signal sequence as predicted by SignalP 3.0 (Bendtsen et al., 2004) with the forward and reverse primers also containing engineered EcoRI and XbaI sites, respectively. PCR reactions were performed using the following cycling conditions: initial denaturing at 94 °C for 2 min; 30 cycles of 94 °C for 20 s, 48 °C for 20 s, and 72 °C for 30 s; and a final extension at 72 °C for 3 min. Amplified cDNA was cloned into pCR2.1-TOPO and transformed into *Escherichia coli* TOP10 according to the manufacturer's protocol (Invitrogen; Carlsbad, CA). Individual clones were isolated and sequenced to identify the intron positions and confirm the fidelity of the sequence.

To isolate the cDNA of genes not obtained by reverse transcription of total RNA, the full-length putative PG genes were amplified using *Pfu* ultra II Fusion HS DNA polymerase (Stratagene; LaJolla, CA) with *R. oryzae* 99–880 genomic DNA as the template. Gene specific forward and reverse primers (Supplementary Table 2) engineered with SphI and NotI restriction sites, respectively, were used to facilitate cloning into the *R. oryzae* expression vector pPdcAEx-6XHIS; a derivative of pPdcAEx (Mertens et al., 2006) containing a multiple cloning site (MCS) and an enterokinase (EK) site and 6X His-tag downstream of the MCS. The forward primers included the native start codons of the PG genes and the reverse primers are designed so the PG genes are in frame with the EK site and 6X His-tag of the expression vector. The PCR reactions were performed using the following cycling conditions: initial denaturing at 94 °C for 2 min; 30 cycles of 94 °C for 20 s, 48 °C for 20 s, and 72 °C for 30 s; and a final extension at 72 °C for 3 min. The amplified genomic DNA was cloned into pCR2.1-TOPO and transformed into *E. coli* TOP10. The individual clones were isolated and sequenced to confirm the fidelity of the sequence. The genomic PG clones were then inserted between the SphI and NotI sites of the pPdcAEx-6XHIS vector and transformed into *E. coli*. The individual clones were isolated and sequenced to confirm the fidelity of the sequence.

Transformation of the *R. oryzae* strain *pyr 17* was performed as previously described (Skory, 2002). Approximately 5–17 days after bombardment, spores were harvested using sterile water and used to inoculate shake flask cultures. Total RNA was isolated as described above from *pyr17* mycelia grown in 50-ml shake flask cultures for 24 h at 30 °C and at 200 rpm in RZ1 minimal medium containing 0.5% glucose. Total RNA was reverse transcribed and first strand synthesis was performed as described above. The cDNA was amplified using gene specific forward primers (Supplementary Table 2) and the reverse primer EKHSinMCSrevshort: 5'-TTAATT AATGATGATGATG-3' to eliminate possible isolation of a similar cDNA from the *pyr17* strain. The cDNA products were cloned into pCR2.1-TOPO and transformed into *E. coli*. Three to five individual clones for each PG construct were sequenced to identify intron positions and confirm the fidelity of the cDNA sequence.

### 2.3. Phylogenetic analyses

Phylogenetic analyses of the polypeptide sequences including all eighteen 99–880 putative PGs and *rpg1* (PG from different strain of *R. oryzae*) along with several additional gene sequences representing various species, including three bacterial PG gene sequences that were used as outgroup taxa (proteins and GenBank accession numbers are listed in the legend of Fig. 4). The polypeptide sequences were aligned using ClustalX v 2.0 (Thompson et al., 1997). Phylogenetic analyses were conducted using the neighbor-

joining (NJ) method implemented in MEGA 4.0 (Tamura et al., 2007). Gamma-corrected distances were generated using the Jones-Taylor-Thornton (Jones et al., 1992) model; the alpha parameter value that was used (2.731) was estimated using the method of Gu and Zhang (1997). Reliability of internal branches was assessed using 1000 bootstrap pseudoreplicates. Bayesian phylogenetic analysis was conducted using the program Mr.Bayes 3.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). An initial run of 100,000 generations was used to estimate the fixed-rate amino acid model used for the analysis. The fixed rate model identified was the Whelan and Goldman (2001) (WAG) model, which was then used to run subsequent analyses. A total of 510,000 generations were run with a sample frequency of 10 and a burn-in of 25%. The analysis was allowed to proceed until the standard deviation of split frequencies reached 0.002, at which point the analysis was terminated and the results tabulated.

#### 2.4. Expression of PGs in *P. pastoris*

The PG cDNA constructs obtained through RT-PCR from total RNA isolated from strain 99-880 were digested with EcoRI and XbaI and the resulting fragment was ligated into pPICZαA for secreted expression of the putative enzymes in *P. pastoris*. The cDNA constructs created by expression in *R. oryzae* *pyr17* were used as template for PCR using *Pfu* ultra II Fusion HS DNA polymerase and the gene specific forward (EcoRI) and reverse (XbaI) primers (Supplementary Table 1). The resulting product was digested with EcoRI and XbaI and ligated directly into pPICZαA. After sequencing to confirm the fidelity of the constructs, the expression constructs were linearized with PmeI and transformed into *P. pastoris* strain X-33 (Invitrogen; Carlsbad, CA) using electroporation according to the manufacturer's protocol. Transformants were cultured on YPDS + Zeocin plates (1% [wt/vol] yeast extract, 2% [wt/vol] peptone, 2% [wt/vol] dextrose, 1 M sorbitol, 100 µg/ml Zeocin) at 30 °C. After 3 days, 10 isolates for each construct were spotted onto plates containing BMM induction medium (2.8% [wt/vol] yeast nitrogen base with ammonium sulfate, 100 mM potassium phosphate, pH 6.0, 4 × 10<sup>−5</sup> biotin, 1% methanol) and 1.5% pectin and grown between 48 and 96 h for initial activity screening. The plates were washed with deionized water to remove the cells and then flooded with a 1% CTAB [wt/vol] solution for 1 h (Hankin et al., 1971). The plates were rinsed with deionized water and then stained with a 0.1% ruthenium red [wt/vol] solution for 2 h. Pectin hydrolysis was detected as halos around the colonies. The most active isolates for each construct, as determined by the pectin plate analysis, were cultured overnight in 3 mls YPD at 30 °C and 225 rpm. The cells were transferred to a 250-ml baffled flask containing 50 mls BMG medium (2.8% [wt/vol] yeast nitrogen base with ammonium sulfate, 100 mM potassium phosphate, pH 6.0, 4 × 10<sup>−5</sup> biotin, 1% glycerol) and grown overnight to increase cell density. The cells were harvested, washed, and transferred to a baffled flask containing 50 mls BMM induction medium and grown for 48 h (1% MeOH was added after 24 h to maintain induction). The cell-free supernatants were concentrated using the Amicon Ultra-15, PLBC Ultracel-PL Membrane with a 3 KDa cutoff (Millipore; Billerica, MA) and buffer exchanged with 50 mM sodium acetate, pH 5.0. The protein content of the concentrated supernatant was assayed by the method of Bradford (BioRad; Hercules, CA) with bovine serum albumin as the standard. The samples were frozen at −20 °C until further use.

#### 2.5. Enzyme assays

PG activity of the concentrated cell culture supernatants was performed using the Nelson-Somogyi assay (Somogyi, 1952), with PGA as the substrate. Briefly, a 500 µl aliquot of the supernatant (48 h

induction media) was mixed with 2.5 ml 0.3% PGA in 75 mM sodium acetate, pH 5.3, and incubated at 30 °C. At various time points, 500 µl aliquots of the reaction mixtures were combined with 500 µl of the copper reagent and immediately boiled for 10 min. The insoluble reaction components were removed by centrifugation and the absorbance of the supernatant was read at 500 nm. D-Galacturonic acid sodium salt was used as the standard. Analysis of the PGA hydrolysis reaction products was performed using thin layer chromatography (TLC). For each time point of the reactions performed above, a 3 µl aliquot was spotted onto Whatman Partisil HPLC TLC Plates (Silica Gel 60 Å, 10 × 10 cm, 200 µm thick; GE Healthcare; Piscataway, NJ). The mobile phase consisted of a 5:3:2 mixture of *n*-butanol:H<sub>2</sub>O:acetic acid (Lojkowska et al., 1995). The products were visualized by quickly soaking the plates in a 0.2% orcinol [wt/vol in methanol] and 10% sulfuric acid solution, followed by a short heating (Cabanne and Doneche, 2002). Standards consisted of 1 µg each mono-, di-, and tri-galacturonic acid.

### 3. Results and discussion

A search of the recently sequenced *R. oryzae* strain 99-880 genome database uncovered the presence of eighteen putative PG genes that we have subsequently identified as *rpg1*–*18*. The putative genes range from ~1.2 kb to 1.35 kb in size, with only one of the putative genes nearly identical to the previously reported *rpg1* gene from *R. oryzae* strain YM9901 (Yoshida et al., 2004). Two of the genes, *rpg2* and *rpg3*, are identical and found consecutively on the same chromosome, albeit on opposite strands. Aside from the identical *rpg2* and *rpg3* genes, the genes share from 50% to greater than 90% homology based on Lipman-Pearson methods (Lipman and Pearson, 1985), with the majority of divergence occurring in the number and/or length of the predicted introns. The *rpg* genes are predicted to contain anywhere from one to three introns, and all possess a first intron that is found in a nearly identical position (Table 1). The position of the predicted 2nd and 3rd introns, when present, follows the same pattern.

The presence of large families of pectinolytic genes in fungal species is not uncommon. However, the eighteen *R. oryzae* PG genes represent an unusually large gene repertoire for a single genome. For example, the ascomycete fungal species *A. niger*, has eleven PG genes and the basidiomycete species *B. cinerea* has six genes (Pel et al., 2007; Kars et al., 2005). The most notable difference among the *R. oryzae* PG genes, much like the *A. niger* PG genes, is the number and length of the introns. A number of *rpg* genes are also in close proximity. The genes encoding *rpg2* through *rpg4* are within a 15 kb region on chromosome 10 and all contain one intron with *rpg2* and *rpg3* found consecutively and likely the result of a recent duplication event (Table 1). Similarly, *rpg14*–*16* are within a 40 kb region on chromosome 4 with *rpg14* and *rpg15* found consecutively and all three genes containing three introns. In addition, unlike the other *rpg* genes, these three genes also demonstrate a great deal of sequence similarity up to 150 bp upstream of the start codon (data not shown).

The nucleotide sequence 300 bp upstream of the putative start codon of the other 15 *rpg* genes does not demonstrate a high degree of sequence similarity, with perhaps the exception of the location of the putative transcription start sites (data not shown). The TATA box is located at or near −76 bp for all 18 genes. The number and location of the CAAT motif is highly variable among the genes in the region upstream of the start codon and they lack promoter sequence motifs suggested to play a role in transcriptional activation of Ascomycete and Basidiomycete PG genes (Bussink et al. 1992, deVries and Visser, 2001; Williams et al. 2002). The 3' flanking ends of the 18 PG genes also exhibits low sequence similarity; however, all possess an AATAAA polyadenylation signal anywhere from 29 to 187 nt downstream of the stop codon.

**Table 1***Rhizopus oryzae* 99–880 polygalacturonase genes.

Gene name	Locus <sup>*</sup> number	Chromosome <sup>*</sup>	Sequence location <sup>*</sup> bp (Strand)	Introns/location	cDNA bp	Protein AA mature
<i>rpg1</i>	04731-1	7	1404558–1405758(–)		1152	356
<i>rpg2</i>	15510-1	10	614845–616074(–)		1152	356
<i>rpg3</i>	15511-1	10	618377–619606(+)		1152	356
<i>rpg4</i>	15419-1	10	363664–364867(+)		1152	357
<i>rpg5</i>	15508-1	10	605229–606459(–)		1152	356
<i>rpg6</i>	11297-1	9	542467–543733(+)		1158	360
<i>rpg7</i>	02738-1	2	1335777–1337105(+)		1152	357
<i>rpg8</i>	15015-1	15	579781–581090(+)		1149	356
<i>rpg9</i>	07548-1	5	2058672–2059876(+)		1146	355
<i>rpg10</i>	08338-1	3	888447–889707(–)		1146	355
<i>rpg11</i>	06273-1	4	2103345–2104607(–)		1149	356
<i>rpg12</i>	08126-1	3	390787–392110(+)		1146	355
<i>rpg13</i>	02911-1	2	1786115–1787423(–)		1155	358
<i>rpg14</i>	06020-1	4	1420993–142272(+)		1158	359
<i>rpg15</i>	06012-1	4	1381338–1382658(–)		1158	359
<i>rpg16</i>	06021-1	4	1424035–1425382(+)		1158	359
<i>rpg17</i>	05849-1	4	879817–881228(+)		1131	354
<i>rpg18</i>	08304-1	3	820878–822132(–)		1134	355

<sup>\*</sup> From *Rhizopus* Genome Database (*Rhizopus oryzae* Sequencing Project, Broad Institute of Harvard and MIT (<http://www.broad.mit.edu>)). Locus numbers are prefaced by RO3G\_.

The 18 putative PG genes encode for proteins that range in size from 377 to 385 amino acids (Fig. 1), however, the mature peptides are expected to be smaller. The PG peptides are predicted to have signal sequences of 23–27 amino acids (Fig. 1; Bendtsen et al., 2004) resulting in mature peptides of 354–360 amino acids leading to molecular weights of roughly 38 kD (Table 1). The predicted peptides all possess the highly conserved amino acids previously demonstrated to be required for enzymatic activity in previously isolated fungal PG enzymes (Fig. 1). The putative PG's contain the three conserved aspartic acid residues found in the active site of the enzyme, the consecutive aspartic acid residues D208 and D209 (RPG1 numbering), along with D187 that is part of the strictly conserved sequence NTD (Fig. 1). The putative enzymes also possess the conserved residues, H231, thought to be important in maintaining the proper ionization state of the active site and R265, K267, and Y299, residues implicated in substrate binding (van Santen et al., 1999; Pagès et al., 2000). Crystal structures of *A. niger* PG's have shown the existence of four disulfide bridges (van Santen et al., 1999; Cho et al., 2001) and this feature appears to be conserved in the putative *R. oryzae* PG enzymes as there are eight conserved cysteine residues in positions comparable to the *A. niger* enzymes. In addition, the *R. oryzae* proteins possess four additional cysteine residues conserved in all 18 putative enzymes suggesting the possibility of additional disulfide bridges.

### 3.1. Isolation of *R. oryzae* PG cDNA

Since *rpg1* was the only PG transcript that had been isolated previously from *R. oryzae* (Yoshida et al., 2004), we questioned whether the additional genes were actively transcribed. To gain some understanding on the expression of the identified *R. oryzae* PG genes and isolate a cDNA copy for each gene to enable recombinant protein production, RNA was isolated from mycelia of submerged and solid *R. oryzae* 99–880 cultures grown using multiple carbon sources. Despite the multiple growth conditions, only 6 of the 17 cDNA sequences were obtained. The cDNA's of *rpg1*, *rpg12*, and *rpg14* were isolated from cultures with pectin as the carbon

source. Two cDNA's, *rpg15* and *rpg16*, were isolated in cultures containing D-galacturonic acid, while the *rpg 18* cDNA was isolated using sucrose as the carbon source. The isolated cDNA clones for the six genes demonstrated the predicted intron number and splice junctions leading to the predicted cDNA sequence. While the possibility exists that a number of the identified genes are not transcribed due to divergence in the promoter region, it is also possible that a number of these genes are expressed at a very low level and below the limits of detection in our assay. The attempts at direct isolation of the cDNA for each gene cannot be considered exhaustive even though a number of clones obtained for each RT-PCR reaction were sequenced due to the high sequence similarity among the genes. It follows too, that the cDNA's we were able to isolate under the stated conditions may also be expressed in cultures using additional carbon sources.

The three genes expressed when pectin was used as the carbon source, *rpg1*, *rpg12* and *rpg14*; as will be discussed in detail below, encode active endo-PG (*rpg1*, *rpg12*) and an exo-PG (*rpg14*). Considering the role endo-PG's and exo-PG's play in the breakdown of pectin it is not unexpected that both endo- and exo-PG enzymes would be expressed simultaneously. The endo-PG's hydrolyze randomly within the oligogalacturonate chain, while the exo-PG's hydrolyze either di- or mono-galacturonic acid from non-reducing ends allowing for complete hydrolysis of the polygalacturonic acid polymer (Markovič and Janaček, 2001; Niture, 2008). The simultaneous expression of endo- and exo-PG's appears to be common as it has been reported that *Aspergillus* species also express both enzyme forms on various substrates including pectin (Kester and Visser, 1990; deVries and Visser, 2001). Furthermore, endo-PG enzymes appear to be required for the initial stage of infection in plants as the exo-PG enzymes tested to date are unable to macerate plant tissue independently and appear to be more important to the latter stages of pectin hydrolysis (Abbott and Boraston, 2007). It is perhaps somewhat unexpected that only *rpg15* and *rpg16*, both exo-PG's (discussed below) were expressed when galacturonic acid was the carbon source as it has been suggested that galacturonic acid acts as a general inducer of pectinolytic genes

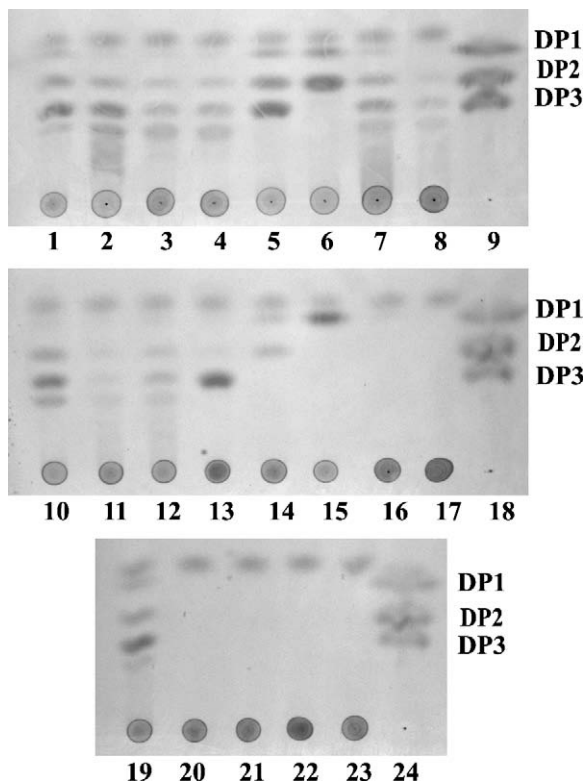


rp1	MVQFLSLSSSSVAFFLLSLCINDVAAAN--C	VVASSSGDAAANIVSAFNSC	NNGGT	VTF	TGK	VTY	NLS	VVT	I	Q	G	L	K	H	V	78					
rp2	MVQFLSLTSSVAALLLLSLGVNDVAAAN--C	VVASSSGDAAANIVSAFNSC	NNGGT	VTF	TGK	VTY	NLS	VVT	I	Q	G	L	K	H	V	78					
rp3	MVQFLSLTSSVAALLLLSLGVNDVAAAN--C	VVASSSGDAAANIVSAFNSC	NNGGT	VTF	TGK	VTY	NLS	VVT	I	Q	G	L	K	H	V	78					
rp4	MVQFLSLTSSVAALLLLSLGINNVAAATE-	CVVARSGGDANNIIISAFNCK	NGGT	VTF	TGK	VTY	NLS	SMIE	IT	GL	KN	V				78					
rp5	MVQFLSLTSSVAALLLLSLGINDVAAAGN-	CVVRSRGDDANNIIISAFNCK	NGGT	VTF	TGK	VTY	NLS	SMIE	IT	GL	KN	V				78					
rp6	MVQFLSLSSSSVAFFLLSLGVNNVAGATN	NCVVSASGSDDTSAITEAFKK	CASGGT	VTF	SKG	STY	NK	D	S	V	T	E	GL	KN	V	80					
rp7	MVQFLSLTSSVAALLLLSLGVNVASATTT-	CAVASSSGDAADSNIVTAFT	CNNNGGT	VTF	TGK	VTY	NLS	IVT	IT	GL	KN	V				78					
rp8	MVQFLSLTSSIAALLLLSAGISDVSAATTT-	CVVASSSGDAAANIVSAFT	CKSGGT	VTF	TGK	VTY	NLS	IVT	IS	GL	KN	V				78					
rp9	MVQLLSLTSSVATLLLLSFGFNEVVAAGTKT-	CTVAKSSDDAIIATIAQAFN	CCKNGGT	VTF	PKG	TK	Y	M	K	S	L	V	S	V	GL	KN	78				
rp10	MVQLLSLSSSIAALLLLVSLGINDVAAAT-	CVVNPTGGDDASAIKAFET	CKSGGT	VTF	PKG	TK	Y	M	K	S	L	V	S	V	GL	KN	77				
rp11	MVQILSLSSSSVAALLLLSLGINDVAAAGST-	CVVKPTGGDDASAIKAFQNC	CKSGGT	VTF	PKG	TK	Y	M	K	S	L	V	S	V	GL	KN	78				
rp12	MVQLLSLSSSIAALLLLVSLGINDVAAAT-	CVVNPTGGDDASAIKAFQNC	CKSGGT	VTF	PKG	TK	Y	M	K	S	L	V	S	V	GL	KN	77				
rp13	MVKFLSLTSSVATLLLLSLGAGVAAATAT-	TCTVAKSGSDDTAAITKAF	SDCKNGGT	VTF	SKD	TTY	N	L	G	I	L	S	D	L	KN	V	79				
rp14	MVRFISLTNISVATLLLLSFRGHTLTAATN	TNTCVVAKDSSDDAIIAKAF	EDCKT	DGT	VTF	V	S	K	N	T	Y	S	L	V	S	GL	KN	80			
rp15	MVRFISFTSPVATALLLLSFGVKHASTAST	NTCIVANSDDSDAIIAIEAF	EKCKT	DGT	VTF	PKD	S	S	Y	Q	L	N	S	I	V	T	S	D	L	KN	80
rp16	MVRFTSTFTSPFSAIILLSPGINKVATATN	TCTCVVAKSDSDDAITLIEAF	EKCKT	DGT	VTF	PKD	S	S	Y	Q	L	N	S	I	V	T	S	D	L	KN	80
rp17	MVQIFSTTSLSLATLLLIASSA---AAATK	T---CVVQNKSDSDSISITQ	AFNDCKNGGT	VHF	PR	G	K	T	Y	P	K	S	L	I	K	S	GL	KN	74		
rp18	MVQFFSASLSFFATLLLLTSAS---AAQKV-	---CVVKN	SRTDSDSITITQAF	NDCKAGGT	VTF	PKN	S	K	Y	N	I	K	S	L	I	K	V	T	GL	KN	75
rp1	NVNLAGTINLTPARSSSFKNLDHYIELKGDNV	QVFGGTVNGNQAWYDAADHTAPT	VFRFSL	TNS	SKVSG	NI	INAPRAH										158				
rp2	NVNLAGTINLTPARSSSFKNLDHYIELKGDNV	QVFGGTVNGNQAWYDAADHTAPT	VFRFSL	TNS	SKVSG	NI	INAPRAH										158				
rp3	NVNLAGTINLTPARSSSFKNLDHYIELKGDNV	QVFGGTVNGNQAWYDAADHTAPT	VFRFSL	TNS	SKVSG	NI	INAPRAH										158				
rp4	NINFAGGINLTPARASFKQGGQNHYYIQLKGD	NVMYGGGTINGNQAWYDAADHTAPT	VIRFALT	TNS	SKVSH	FN	INAPRAH										158				
rp5	NINFAGGINLTPARASFKQGGQNHYYIQLKGD	NVMYGGGTINGNQAWYDAADHTAPT	VIRFALT	TNS	SKVSH	FN	INAPRAH										158				
rp6	NINFAGGINLTPARASFKQGGQNHYYIQLKGD	NVMYGGGTINGNQAWYDAADHTAPT	VIRFALT	TNS	SKVSH	FN	INAPRAH										158				
rp7	NINFAGTLLNTPARSSSFKNLDHYIELKGDNV	MYGGGTINGNQAWYDAADHTAPT	VRFALT	TNS	SKVSG	IK	INAPRAH										160				
rp8	NVNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp9	NLNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp10	NLNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp11	NLNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp12	NLNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp13	NLNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp14	NLNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp15	NLNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp16	NLNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp17	NLNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp18	NLNLAGTIKLARSATSPQNLDHYIQLKGSNV	YVGGGTINGNQAWYDAADHTAPT	VLRLSL	TDS	NVGYFTI	INS	RAHL										158				
rp1	KIDFQGGKITLSPGDKKFKGKNAYIKIRGDN	VHLSGGGTINGNQAWYDLQDHTAPT	VLRLM	MTNS	VFS	NFKI	INAPRAH										157				
rp2	KIDFQGGKITLSPGDKKFKGKNAYIKIRGDN	VHLSGGGTINGNQAWYDLQDHTAPT	VLRLM	MTNS	VFS	NFKI	INAPRAH										157				
rp3	KIDFQGGKITLSPGDKKFKGKNAYIKIRGDN	VHLSGGGTINGNQAWYDLQDHTAPT	VLRLM	MTNS	VFS	NFKI	INAPRAH										157				
rp4	NIELAGTIKLPGFDKAVKGLKSYIQLGNTV	KYVGGVINGNQAWYDAADHTAPT	VLRI	SAN	HSS	VD	SI	INAPRAH									159				
rp5	NVDLAGTIELPARNTSYSTSAFHIEG	SNINLYDGLISG---AWYDAL	DHTAPT	VL	SAN	HSS	VD	SI	INAPRAH								157				
rp6	NINFAGTIHLPARAEESTRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp7	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp8	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp9	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp10	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp11	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp12	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp13	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp14	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp15	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp16	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp17	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp18	NINLAGTINLPVREESYRNGDYIYIQLG	THIKMYGGGTINGNQAWYDALDHTAP	SVLRIA	AND	S	I	IG	FTI	INAPRAH								160				
rp1	GVTNSQNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp2	GVTNSQNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp3	GVTNSQNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp4	GATNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp5	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp6	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp7	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp8	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp9	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp10	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp11	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp12	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp13	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp14	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp15	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp16	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp17	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp18	GVTNSKNVLVDNITLKTAS--TSSKLA	NTDALDVSSSGGI	IFQNSH	LT	V	G	D	D	C	L	A	I	N	G	G	V	237				
rp1	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp2	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp3	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp4	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp5	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp6	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp7	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp8	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp9	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp10	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp11	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			
rp12	LGKGGATENVNMVRI	LSSTC	NTCQNGV	R	I	K	T	W	P	G	G	K	G	V	S	D	V	316			

**Fig. 1.** Alignment of *R. oryzae* RA99–880 deduced PG amino acid sequences RGP1–RGP18. The outlined residues denote conserved regions of family 28 polygalacturonase family members. The asterisks highlight residues required for enzymatic activity. The + symbol above the highlighted amino acids indicates the first amino acid of the expected mature peptide. The ^ symbol highlights the conserved cysteine residues with lines between cysteine residues depicting potential disulfide bridges that are found in similar region of *A. niger* Endo II polygalacturonase enzyme crystal structure (van Santen et al., 1999).

in *A. niger* (deVries and Visser, 2001). This point, along with the lack of discernable transcriptional activator elements in the promoter region, suggests that there are some differences in the regulation of the *Rhizopus* PG genes in comparison to the *Aspergillus* genes. It is possible that the exo-PG's, *rpg15* and *rpg16*, are activated only after the initial stages of a fungal infection when galacturonic acid would be present and function to rapidly hydrolyze remaining oligogalacturonates to mono- and digalacturonates for full utilization of the pectic substrate. This possibility also fits with the suggestion that the large repertoire of PG genes is required to increase efficiency of degradation and to obtain nutrients at different phases of the life cycle (Williams et al., 2002), however; work beyond what is presented here is required to fully develop this possibility.

Although we could not isolate the cDNA for each gene directly using multiple carbon sources, we were able to obtain the remaining cDNA clones by recombinant methods. A genomic clone of the remaining genes was obtained via PCR and confirmed by sequencing. The *rpg* genes were then placed into recombinant expression vector pPdcAEx-6XHIS and subsequently transformed into *R. oryzae* strain *pyr17* for expression (Mertens et al., 2006). Total RNA was isolated and the cDNA for the remaining PG genes isolated via RT-PCR using a gene specific 5' primer and a 3' primer specific to the 6XHis region of the vector to ensure that cDNA's obtained were from the recombinant construct and not the *pyr17* strain used for expression. The 11 remaining cDNA's were obtained by this method and exhibited the predicted intron number and the expected splice junctions for each gene (Table 1 and data not shown).

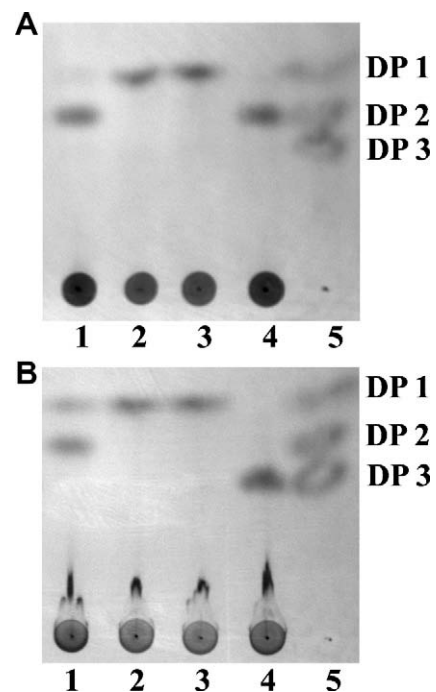


**Fig. 2.** Thin layer chromatography separation of hydrolysis products after a 60 min hydrolysis reaction containing PGA and *P. pastoris*-expressed *R. oryzae* PG extracts (4 µg total protein) as detailed in Section 2. Lanes 9, 18, and 24: Standards; DP1, mono-galacturonic acid; DP2, di-galacturonic acid; DP3 tri-galacturonic acid. Lane 1: RPG1; Lane 2: RPG2; Lane 3: RPG4; Lane 4: RPG5; Lane 5: RPG6; Lane 6: RPG7; Lane 7: RPG8; Lane 8: RPG9. Lanes 10–17: RPG10–RPG17; Lane 19: RPG1; Lane 20: RPG18; Lane 21: *Pichia* X-33 empty pPICZαA vector; Lane 22: *Pichia* X-33 no vector; Lane 23: Buffer blank, no enzyme.

In addition, it should be pointed out that there are two distinct “types” of *R. oryzae* strains denoted Type I, in the case of lactic acid-producing strains, and Type II, in the case of fumaric acid-producing strains. There are clear genetic differences between the two different strain types (Abe et al., 2003; Saito et al., 2004b). Yet, the presence of eighteen PG genes apparently predates their divergence, as we were able to confirm that all eighteen genes present in *R. oryzae* 99–880 (a Type II strain) were also present in *R. oryzae* NRRL 395 (a Type I strain). While transcription of the 18 *rpg* genes is not entirely clear at this time, work to gain additional insight into the expression of these genes is required.

### 3.2. Expression of the putative PG enzymes in *P. pastoris*

Due to the inordinately high number of PG genes detected, the high sequence similarity, and since transcripts for all eighteen *rpg* transcripts could not be detected under the growth conditions used, the putative PG genes were expressed in *P. pastoris* to determine if the genes do in fact code for active PG enzymes. The seventeen *rpg* cDNA's with the putative signal sequence deleted were placed in frame with the α-mating signal of *P. pastoris* vector pPICZα-A for secreted expression (expressed 17 of 18 since *rpg2* and *rpg3* are identical). All seventeen of the expressed genes produced the desired gene product (data not shown) and fifteen of the seventeen expressed proteins demonstrated measurable activity, with no detectable activity arising from the X-33 *Pichia* strain or *Pichia* X-33 transformed with an empty vector. Hydrolysis reactions, with 60 min and 24 h time points, were performed using concentrated culture supernatants and PGA as the substrate followed by separation of the products on thin layer chromatography plates (Fig. 2). The 60 min hydrolysis reactions show that RPG1–13 can be classified as endo-PGs since multiple products ranging from mono-galacturonic acid (DP1) to hexa-galacturonic acid (DP6) could be



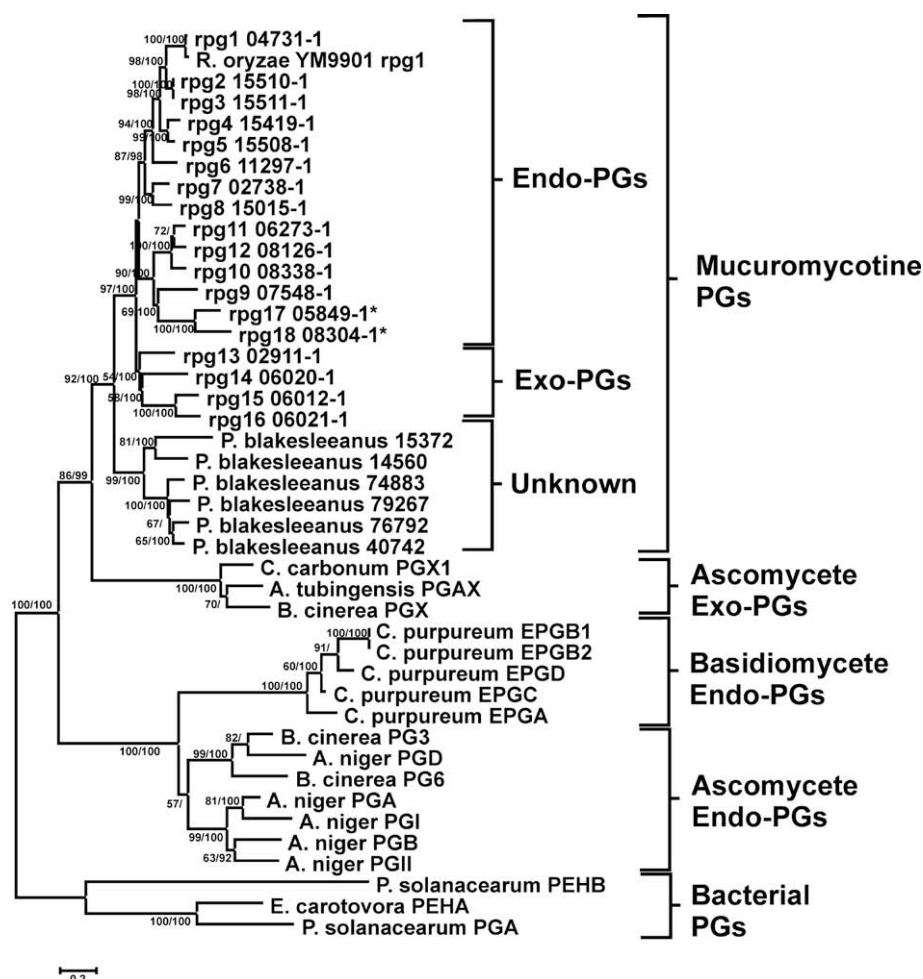
**Fig. 3.** Thin layer chromatography separation of hydrolysis products after a 60 min hydrolysis reaction containing either (A) di-galacturonic acid (DP2) or (B) tri-galacturonic acid (DP3) and *Pichia*-expressed *Rhizopus* polygalacturonase extracts (4 µg total protein) as detailed in Section 2. Samples loaded in identical manner in (A and B). Lane 1: RPG14; Lane 2: RPG15; Lane 3: RPG16; Lane 4: RPG17; Lane 5: Standards—DP1, DP2, and DP3.

detected (Fig. 2 lanes 1–8 and 10–13). In addition, carrying out the reaction to completion resulted in the expected reaction products mono- and di-galacturonic acid; however the RPG1–13 enzymes do not demonstrate activity towards di-galacturonic acid (data not shown). RPG13 demonstrates a different product profile from RPG1–RPG12 in that it has a strong product preference for DP3 with a small amount of DP2 present (Fig. 2, lane 13). RPG7 also demonstrates a slightly different product profile and gives the appearance (Fig. 2, lane 6) of an exo-PG since the weak DP3 band on the TLC plate is not visible in the figure, however, the enzyme does not hydrolyze di-galacturonic acid as would be expected for an exo-PG and hydrolysis of tri-galacturonic acid leads to di- and mono-galacturonic acid products.

The RPG14, 15, and 16 enzymes are exo-PGs as D-galacturonic acid is the main product with RPG14 also having di-galacturonic acid as a minor product (Fig. 2, lanes 14–16). All three were able to cleave di-galacturonic acid, a hallmark of exo-PG's, but RPG14

does not appear nearly as efficient as RPG 15 and 16 (Fig. 3A, lanes 1–3). When DP3 is used as the substrate, D-galacturonic acid is again the only product detected in RPG15 and RPG16 reactions; however, hydrolysis by RPG14 resulted in both di- and mono-galacturonic acid as products (Fig. 3B; lanes 1–3).

The final two *rpg* genes, *rpg17* and *rpg18*, encode proteins that do not lead to any detectable PGA hydrolysis products (Figs. 2 and 3), despite the presence of the conserved amino acids required for catalysis (Fig. 1). Viscosity assays, as well as assays using pectin with varying degrees of methyl esterification and rhamnogalacturonan, also failed to demonstrate any activity by the two putative enzymes (data not shown). There are some minor sequence differences specific to these two proteins, relative to the other RPG enzymes, near the conserved tyrosine residue Y299 (RPG1 numbering). RPG17 and RPG18 have a phenylalanine residue directly after Y299 and a deletion a few residues beyond the additional phenylalanine residue (Fig. 1). Whether these subtle changes impact substrate binding,



**Fig. 4.** Phylogeny of the PG multigene family. Three bacterial PGs were included as outgroup taxa in order to root the tree. Because both NJ and Bayesian trees were highly similar in topology, only the Bayesian tree is shown here. Numbers along branches represent Bayesian posterior probabilities (before the slash mark) and NJ bootstrap values (after the slash mark). Only values greater than 50% are shown. The numbers of amino acid substitutions per site are given to scale. The publication sources and GenBank accession numbers for the amino acid sequences are as follows: *A. niger* PGI (Bussink et al., 1991; X58892), PGII (Bussink et al., 1990; X58893), PGA and PGB (Parenicová et al., 2000a; Y18804 and Y18805), PGD (Parenicová et al., 2000b; Y18806); *A. tubingensis* PGAX (Kester et al., 1996; X99795); *B. cinerea* PG3 and PG6 (Wubben et al., 1999; U68717 and U68722), PGX (AF145229); *C. purpureum* EPGA, EPGB1, EPGB2, EPGC, and EPGD (Williams et al., 2002; AF237653, AF237654, AF348967, AF237655, and AF237656); *Cochliobolus carbonum* PGX1 (Scott-Craig et al., 1998; L48982); *Erwinia carotovora* PEHA (Saarilahti et al., 1990; X51701); *Pseudomonas solanacearum* PGA (Huang and Schell, 1990; M33692) and PEHB (Huang and Allen, 1997; U60106); and *R. oryzae* YM99021 *rpg1* (Yoshida et al., 2004; AB127052). The remaining *R. oryzae* sequences ([http://www.broad.mit.edu/annotation/genome/rhizopus\\_oryzae/MultiHome.html](http://www.broad.mit.edu/annotation/genome/rhizopus_oryzae/MultiHome.html)) and the *Phycomyces blakesleeenans* sequences (<http://genome.jgi-psf.org/Phybl1/Phybl1.home.html>) were extracted from the complete genome databases of those species. The *Phycomyces* sequences are identified by their respective protein ID numbers which are a searchable term in the genome database. In the tree, we dubbed the latter group as "Unknown" because their function has not been assayed. Strains denoted by an asterisk failed to yield a result in the expression analyses; however, because the phylogenetic placement of these genes with other endo-PGs is supported by high posterior probabilities and NJ bootstrap values, we consider them to be a part of the endo-PG phylogenetic cluster and leave them named as such.



since Y299 has been implicated in substrate binding (van Santen et al., 1999; Pagès et al., 2000), or impact the subsite structure is unclear. Interestingly and despite the apparent inactivity, *rpg18* is actively transcribed in the presence of sucrose; however, expression of a seemingly inactive protein is not unprecedented in *R. oryzae*. The *amyB* gene of *R. oryzae* NRRL 395 was shown to be expressed under a number of different culture conditions, yet also failed to demonstrate any activity (Mertens and Skory, 2007). Clearly, additional work will be required before concluding that *rpg17* and *rpg18* are definitively pseudogenes.

### 3.3. Phylogeny of the PG multigene family

Prior to the sequencing of the *R. oryzae* genome, only a single PG gene (*rpg1*) had been identified from *R. oryzae*. In this study, we identified an additional seventeen putative genes, of which fifteen were found to be functionally active. No other species of fungus has been shown to possess such a large number of PG genes, suggesting that *R. oryzae* genes may have arisen subsequent to their divergence from other fungi. To test this hypothesis, we reconstructed the phylogenetic relationships of the PG multigene family using genes from the representative species of Mucoromycotina (*R. oryzae* and *Phycomyces blakesleeanus*), Basidiomycota (*C. purpureum*), and Ascomycota (*A. niger*, *A. tubingensis*, *B. cinerea*, and *Cochliobolus carbonum*) included in this study. The results of our analyses (Fig. 4) show that mucormycotine genes form their own distinct clade, consisting of a subclade of *R. oryzae* endo- and exo-PG genes and a subclade of *P. blakesleeanus* genes whose function has not been assayed, but are clearly homologous to other fungal PG genes included in this study. This result confirms our initial hypothesis that mucormycotine PG genes arose subsequent to this lineage's divergence from the ascomycete and basidiomycete genes. Otherwise, we would have observed instances of mucormycotine genes interspersed with ascomycete and basidiomycete genes. Similarly, the basidiomycete PG genes (all of which were endo-PGs) formed their own separate cluster, indicating that those genes arose after basidiomycetes diverged from ascomycetes. Interestingly, however, the ascomycete genes formed two distinct clusters: an exo-PG gene clade that clustered with mucormycotine PG genes and an endo-PG gene clade that clustered with basidiomycete endo-PG genes. These relationships were supported with high bootstrap and Bayesian posterior probability values (Fig. 4). Furthermore, our results also suggest that the rate of PG gene turnover is elevated. Gene turnover refers to the rate at which new genes arise via gene duplication and are lost as a result of deletion or pseudogenization and deterioration (Rooney and Ward 2005). The tree in Fig. 4 shows that *Rhizopus* genes are distinct from the *Phycomyces* genes, indicating that these subclades arose after *Rhizopus* and *Phycomyces* diverged from one another. It would be interesting to investigate just how fast the rate of PG gene turnover is by examining the phylogenetic relationships of a group of PG genes from closely related sibling species.

In light of the above clustering patterns, one wonders if the ancestral fungal PG was an exo-PG or an endo-PG. We believe that it was an endo-PG and that exo-PG activity evolved at a later time and at least twice independently: once in the mucormycotines and once in the ascomycetes. We base this reasoning on the following. First, we observed a functional “grade” of exo-PG activity in the exo-PG clade. Specifically, gene *rpg13* (which clusters at the base of the clade) displays moderate endo-PG activity and an apparent product preference for DP3, unlike *rpg1-12* that demonstrates multiple hydrolysis products, but does not show any activity towards di-galacturonic acid (Fig. 2). Next, gene *rpg14* (the next gene to diverge) displays exo-PG activity, however; it has weak activity towards di-galacturonic acid. Finally, genes *rpg15* and *16* (which are the most highly derived members of the clade) displays

strong exo-PG activity as demonstrated by the ability to hydrolyze di-galacturonic acid with mono-galacturonic acid as the sole product of galacturonic acid hydrolysis (Figs. 2 and 3). This pattern is consistent with a scenario in which an ancestral mucormycotine endo-PG gene duplicated and one of the duplicates underwent three further duplications to form the exo-PG clade, in which exo-PG activity evolved. The other duplicate underwent a series of further gene duplication events (at least 17) but endo-PG activity was retained in all cases. Similarly, in the first gene duplication event at the base of the fungal PG gene phylogeny, endo-PG activity was retained; later a second duplication event giving rise to the mucormycotine PGs and the ascomycete exo-PG clade, led to exo-PG activity evolving independently in the latter subsequent to their divergence from the mucormycotines.

Of course, our above interpretations concerning the evolution of exo-PG activity from an endo-PG ancestor depend on the underlying assumption that the species chosen are truly representative. A number of putative Ascomycete and Basidiomycete polygalacturonase sequences related to the ones included in our analyses are available from various fungal genome sequencing projects, but have not been biochemically characterized to determine function. As a result, we chose not to include these putative genes in our phylogenetic study. Future work should also focus on the characterization of PG genes from other fungal lineages not included here, such as the Glomeromycetes, Blastocladiomycetes, Chytridiomycetes, and Microsporidia. Examination of the PGs from these taxa may shed additional light on the evolution of the PG multigene family at even deeper levels of divergence.

### Acknowledgment

We thank K. O'Donnell of the ARS Culture Collection for providing us with *R. oryzae* NRRL 395 and C. Skory for providing us with the *R. oryzae* *pyr17* mutant strain.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fgb.2008.09.009.

### References

- Abbott, D.W., Boraston, A.B., 2007. The structural basis for exopolysaccharuronase activity in a family 28 glycoside hydrolase. *J. Mol. Biol.* 268, 1215–1222.
- Abe, A., Sone, T., Sujaya, I.N., Saito, K., Asano, K., Tomita, F., 2003. rDNA ITS sequence of *Rhizopus oryzae*: its application to classification and identification of lactic acid producers. *Biosci. Biotechnol. Biochem.* 67, 1725–1731.
- Bendtsen, J.D., Nielsen, H., vonHeijne, G., Brunak, S., 2004. Improved prediction signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795.
- Bussink, H.J., Kester, H.C., Visser, J., 1990. Molecular cloning, nucleotide sequence and expression of the gene encoding prepro-polygalacturonase of *Aspergillus niger*. *FEBS Lett.* 273, 127–130.
- Bussink, H.J.D., Brouwer, K.B., de Graaff, L.H., Kester, H.C.M., Visser, J., 1991. Identification and characterization of a second polygalacturonase gene of *Aspergillus niger*. *Curr. Genet.* 20, 301–307.
- Bussink, H.J.D., Buxton, F.P., Fraaye, B.A., de Graaff, L.H., Visser, J., 1992. The polygalacturonases of *Aspergillus niger* are encoded by a family of diverged genes. *Eur. J. Biochem.* 208, 83–90.
- Cabanne, C., Doneche, B., 2002. Purification and characterization of two isozymes of polygalacturonase from *Botrytis cinerea*. *Microbiol. Res.* 16, 1183–1195.
- Cho, S.W., Lee, S., Shin, W., 2001. The X-ray structure of *Aspergillus aculeatus* polygalacturonase and a modeled structure of the polygalacturonase-octagalacturonate complex. *J. Mol. Biol.* 314, 863–878.
- de Vries, R.P., Visser, J., 2001. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiol. Mol. Biol. Rev.* 65, 497–522.
- Gu, X., Zhang, J., 1997. A simple method for estimating the parameter of substitution rate variation among sites. *Mol. Biol. Evol.* 14, 1106–1113.
- Hankin, L., Zucker, M., Sands, D.C., 1971. Improved solid medium for the detection and enumeration of pectinolytic bacteria. *Appl. Microbiol.* 22, 205–209.
- Henriksson, G., Akin, D.E., Slomczynski, D., Eriksson, K.-E.L., 1999. Production of highly efficient enzymes for flax retting by *Rhizomucor pusillus*. *J. Biotechnol.* 68, 115–123.



- Henrissat, B., Davies, G., 1997. Structural and sequence-based classification of glycosyl hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644.
- Huang, J.H., Schell, M.A., 1990. DNA sequence analysis of *pglA* and mechanism of export of its polygalacturonase product from *Pseudomonas solanacearum*. *J. Bacteriol.* 172, 3879–3887.
- Huang, Q., Allen, C., 1997. An  $\alpha$ -D-galacturonosidase, PehB, is required for wild-type virulence of *Ralstonia solanacearum*. *J. Bacteriol.* 179, 7369–7378.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.
- Ishida, H., Hata, Y., Ichikawa, E., Kawato, A., Suginami, K., Imayasu, S., 1998. Regulation of the glucoamylase-encoding gene (*glbB*), expressed in solid-state culture (Koji) of *Aspergillus oryzae*. *J. Ferm. Bioeng.* 86, 301–307.
- Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282.
- Kars, I., Krooshof, G.H., Wagemakers, L., Joosten, R., Benen, J.A.E., van Kan, J.A.L., 2005. Necrotizing activity of five *Botrytis cinerea* endopolygalacturonases produced in *Pichia pastoris*. *Plant J.* 43, 213–225.
- Kester, H.C.M., Visser, J., 1990. Purification and characterization of polygalacturonases produced by the hyphal fungus *Aspergillus niger*. *Biotechnol. Appl. Biochem.* 12, 150–160.
- Kester, H.C.M., Kusters-van Someren, M.A., Muller, Y., Visser, J., 1996. Primary structure and characterization of an exopolygalacturonase from *Aspergillus tubingensis*. *Eur. J. Biochem.* 240, 738–746.
- Lipman, D.J., Pearson, W.R., 1985. Rapid and sensitive protein similarity searches. *Science* 227, 1435–1441.
- Lojkowska, E., Masclaux, C., Boccara, M., Robert-Baudouy, J., Hugouvieux-Cotte-Pattat, N., 1995. Characterization of the *pell* gene encoding a novel pectate lyase of *Erwinia chrysanthemi* 3937. *Mol. Microbiol.* 16, 1183–1195.
- Markovič, O., Janaček, Š., 2001. Pectin degrading glycoside hydrolases of family 28: sequence-structural features specificities and evolution. *Prot. Eng.* 14, 615–631.
- Mertens, J.A., Skory, C.D., Ibrahim, A.S., 2006. Plasmids for expression of heterologous proteins in *Rhizopus oryzae*. *Arch. Microbiol.* 186, 41–50.
- Mertens, J.A., Skory, C.D., 2007. Isolation and characterization of a second glucoamylase gene without a starch binding domain from *Rhizopus oryzae*. *Enzyme Microbiol. Technol.* 40, 874–880.
- Niture, S.K., 2008. Comparative biochemical and structural characterizations of fungal polygalacturonases. *Biologia* 63, 1–19.
- Pagès, S., Heijne, W.H.M., Kester, H.C.M., Visser, J., Benen, J.A.E., 2000. Subsite mapping of *Aspergillus niger* endopolygalacturonase II by site-directed mutagenesis. *J. Biol. Chem.* 275, 29348–29353.
- Parenticová, L., Benen, J.A.E., Kester, H.C.M., Visser, J., 2000a. *pgaA* and *pgaB* encode two constitutively expressed endopolygalacturonases of *Aspergillus niger*. *Biochem. J.* 345, 637–644.
- Parenticová, L., Kester, H.C.M., Benen, J.A., Visser, J., 2000b. Characterization of a novel endopolygalacturonase from *Aspergillus niger* with unique kinetic properties. *FEBS Lett.* 467, 333–336.
- Pel, H.J., de Winde, J.H., Archer, D.B., Dyer, P.S., Hofmann, G., Schaap, P.J., Turner, G., De Vries, R.P., Albang, R., Albermann, K., Andersen, M.R., Bendtsen, J.D., Benen, J.A., van den Berg, M., Breestraat, S., Caddick, M.X., Contreras, R., Cornell, M., Coutinho, P.M., Danchin, E.G., Debets, A.J., Dekker, P., van Dijk, P.W., van Dijk, A., Dijkhuizen, L., Driessen, A.J., d'Enfert, C., Geysens, S., Goosen, C., Groot, G.S., De Groot, P.W., Guillemette, T., Henrissat, B., Herweijer, M., van den Hombergh, J.P., van den Hondel, C.A., van der Heijden, R.T., van der Kaaij, R.M., Klis, F.M., Kools, H.J., Kubicek, C.P., van Kuyk, P.A., Lauber, J., Lu, X., van der Maarel, M.J., Meulenbergh, R., Menke, H., Mortimer, M.A., Nielsen, J., Oliver, S.G., Olsthoorn, M., Pal, K., van Peij, N.N., Ram, A.F., Rinas, U., Roubos, J.A., Sagt, C.M., Schmoll, M., Sun, J., Ussery, D., Varga, J., Vervecken, W., van de Vondervoort, P.J., Wedler, H., Wosten, H.A., Zeng, A.P., van Ooyen, A.J., Visser, J., Stam, H., 2007. Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat. Biotechnol.* 25, 221–231.
- Rhizopus oryzae* Sequencing Project. Broad Institute of Harvard and MIT. <http://www.broad.mit.edu> (accessed December 2006).
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Rooney, A.P., Ward, T.J., 2005. Evolution of a large ribosomal RNA multigene family in filamentous fungi: birth-and-death of a concerted evolution paradigm. *Proc. Natl. Acad. Sci. USA* 102, 5084–5089.
- Saarihtinen, H.T., Heino, P., Pakkanen, R., Kalkkinen, N., Palva, I., Palva, E.T., 1990. Structural analysis of the *pehA* gene and characterization of its protein product, endopolygalacturonase, of *Erwinia carotovora* subspecies *carotovora*. *Mol. Microbiol.* 4, 1037–1044.
- Saito, K., Takakuwa, N., Oda, Y., 2004a. Purification of the extracellular pectinolytic enzyme from the fungus *Rhizopus oryzae* NBRC 4707. *Microbiol. Res.* 159, 83–86.
- Saito, K., Saito, A., Ohnishi, M., Oda, Y., 2004b. Genetic diversity in *Rhizopus oryzae* strains as revealed by the sequence of lactate dehydrogenase genes. *Arch. Microbiol.* 182, 30–36.
- Scott-Craig, J.S., Cheng, Y.Q., Cervone, F., Lorenzo, G., Pitkin, J.W., Walton, J.D., 1998. Targeted mutants of *Cochliobolus carbonum* lacking the two major extracellular polygalacturonases. *Appl. Environ. Microbiol.* 64, 1497–1503.
- Skory, C.D., 2002. Homologous recombination and double strand break repair in the transformation of *Rhizopus oryzae*. *Mol. Genet. Genomics* 268, 397–406.
- Somogyi, M., 1952. Notes on sugar determination. *J. Biol. Chem.* 195, 19–23.
- Stotz, H.U., Bishop, J.G., Bergmann, C.W., Koch, M., Albersheim, P., Darvill, A.G., Labavitch, J.M., 2000. Identification of target amino acids that affect interactions of fungal polygalacturonases and their plant inhibitors. *Physiol. Mol. Plant Pathol.* 56, 117–130.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- van Pouderooyen, G., Snijder, H.J., Benen, A.E., Dijkstra, B.W., 2003. Structural insights into the processivity of endopolygalacturonase I from *Aspergillus niger*. *FEBS Lett.* 554, 462–466.
- van Santen, Y., Benen, J.A.E., Schröter, K., Kalk, K.H., Armand, S., Visser, J., Dijkstra, B.W., 1999. 1.68-Å crystal structure of endopolygalacturonase II from *Aspergillus niger* and identification of active site residues by site-directed mutagenesis. *J. Biol. Chem.* 274, 30474–30480.
- Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699.
- Williams, H.L., Tang, Y., Hintz, W.E., 2002. Endopolygalacturonase is encoded by a multigene family in the basidiomycete *Chondrostereum purpureum*. *Fungal Genet. Biol.* 36, 71–83.
- Wubben, J.P., Mulder, W., ten Have, A., van Kan, J.A.L., Visser, J., 1999. Cloning and partial characterization of endopolygalacturonase genes from *Botrytis cinerea*. *Appl. Environ. Microbiol.* 65, 1596–1602.
- Yoshida, S., Suzuki, F., Tsukiboshi, T., Shinohara, H., 2004. Cloning and characterization of a gene *rpg1* encoding polygalacturonase of *Rhizopus oryzae*. *Mycol. Res.* 108, 1407–1414.
- Zhang, J., Henriksson, H., Szabo, I.J., Henriksson, G., Johansson, G., 2005. The active component in the flax-retting system of the zygomycete *Rhizopus oryzae* sb is a family 28 polygalacturonase. *J. Ind. Microbiol. Biotechnol.* 32, 431–438.